United States Department of Agriculture

Office of the Chief Information Officer

Network Engineering Division

Telecommunications Enterprise Network Design

Application Level Network Traffic Study

Task V Report

May 8, 1998

Executive Summary

Using the Geographic Network Analysis Process (GNAP), the Network Engineering Division has developed a detailed description of the USDA Data Networks. The Network Baseline description, a “snapshot” of the existing USDA Data Networks, required the evaluation of existing network equipment, services, utilization, performance, cost, and survivability. The Network Application Level Traffic Study provides the final portion of the baseline description. The findings of this report re-enforce the conclusions of previous baseline studies.

The Network Application Level Traffic Study examines the USDA Data Network characteristics in terms of network applications. The importance of measuring and correlating network applications with the operational status of the network lies in the consideration of Agency mission programs. The applicability, or “fit”, of a data network for an Agency is ultimately dependent on the operating characteristics of the applications necessary to meet Agency mission program goals.

Key Concepts

- The predominant use of the current USDA Data Networks is World Wide Web browsing on the Internet
- USDA Data Center traffic is currently not a large percentage of the existing IP networks traffic.
- Performance on the USDA Data Network is generally poor due to constraints imposed by existing design topologies.
- Survivability of the current USDA Data Networks is poor due to constraints of the existing design topologies.
- There is now a sufficient description of the present USDA Data Networks available to design TEN alternatives for management evaluation.
- The TEN baseline description indicates that the USDA Data Networks can be significantly improved by a redesign of the existing networks using a more robust design strategy.
- Redesign of the combined USDA Networks will have the biggest impact on cost.

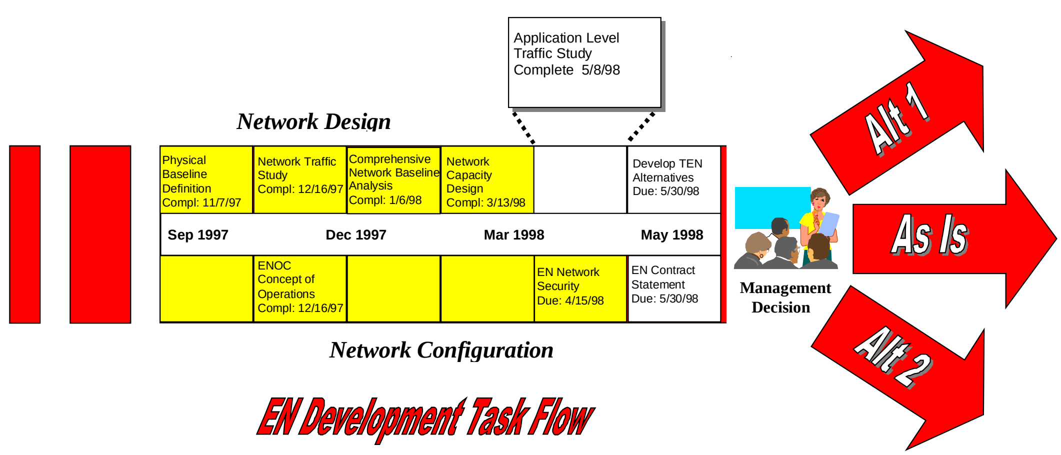
Network Application Study Conclusions

Network Traffic Distribution

- Internet is the predominant use for the existing USDA Data Networks. Of the Internet traffic, Web browsing is the application most in use. Survey data suggests that the percentage of Intranet traffic will increase. For example, there was very little traffic measured on the existing network that went to USDA Data Centers. However, survey data indicates that most agencies do or will communicate with the data centers over the TEN. The conclusion is that there is still traffic going to the data centers on non-IP based networks, such as dial X.25. The continued increase in the X.25 cost in USDA billing data provides further impetus for USDA Agencies to move toward IP based network usage, increasing Intranet traffic on the TEN.

Performance and Survivability

- Topological constraints on critical USDA Data Networks affect performance, including delay, utilization and survivability. Predominant network designs on USDA networks are Minimum Spanning Tree (MST) and hub-and-spoke. Both designs are typically characterized by bottlenecks. Analysis shows that on one USDA network with a design based on a MST structure, the simulated link utilization is 284%. In contrast, a USDA network with a design based on a semi-chordal Ring (SCR) structure has a simulated utilization of 3%. A redesign of the USDA Data Networks will effectively eliminate bottlenecks and improve the overall performance of network applications.
- In general, the performance of the combined USDA data networks is relatively poor. Over-utilization was modeled to be as much as 280% of the theoretical capacity in some circuits. Over-utilization causes transmission bottlenecks that result in infinite delays when simulating a network. In the simulated network, 81.6% of the modeled traffic demands failed due to infinite delays.
- Once the bottlenecks in the model are corrected by increasing bandwidth, the remainder of the networks is basically underutilized. In the modeled network, over 99% of the available bandwidth is not utilized, giving the remainder of the USDA simulated networks very good performance ratings. The majority of the traffic demands delays are in the 0 to 60 msec range indicating very good network performance. As a comparison benchmark, commercial Frame Relay service providers have performance guarantees varying from 70 to 300 msec for delays between the origin and destination POPs of a PVC.
- Network survivability analysis indicates that, on average, single link failures cause a small number of traffic demands to fail. This is due to the large number of “feeder” circuits and LANs associated with small offices. However, there are bottlenecks in the networks that, when failed, cause large numbers of traffic demands to fail. This problem is related to the existing topology design discussed above.



1.0 Introduction

1.1 Objective

The objective of the Network Application Level Traffic Study is to determine the type and volume of network applications used on the present USDA Data Networks. In addition, the study assesses network performance and survivability and makes appropriate recommendations for network optimization. By definition, the study provides the application characteristics baseline to which the Telecommunications Enterprise Network (TEN) design proposals are compared.

1.2 Background

1.2.1 Network Traffic Studies

Traffic studies of the USDA Network are performed at two levels. Total network traffic measurement is necessary to assess the efficiency of current network utilization patterns and trends. In the *Telecommunications Enterprise Network Design, Network Level Traffic Study of USDA Networks: Task II Report, December 12, 1997*, current USDA Data Network usage is shown to be very inefficient relative to available circuit bandwidth. In many cases a recommendation to downsize the circuit results in a cost savings without degradation of performance. However, there is a second level of traffic measurement that must be considered for accurate design recommendations. This second level represents the type, quantity, and business specifications of applications in use by USDA Agencies.

Network application level traffic measurement defines what network applications are in use and how much resource they consume relative to mission program specifications. Network applications include Hypertext Transfer Protocol (HTTP), Transmission Control Protocol (TCP), Network Basic Input/Output System (NetBIOS), Simple Mail Transfer Protocol (SMTP), User Datagram Protocol (UDP), Post Office Protocol version 3 (POP3), Telnet, File Transfer Protocol (FTP), and other TCP/IP traffic associated with Domain Name System (DNS), X-terminal (X11), Trivial File Transfer Protocol (TFTP) etc. Network engineering design requires information about total circuit utilization and capacity as well as performance requirements of specific USDA Agency mission applications.

1.2.2 Network Performance and Survivability

Telecommunications network design efforts attempt to optimize circuits in terms of cost and performance. One of the objectives of network design is cost reduction; equally important objectives are network performance and survivability. Optimizing network performance is critical to USDA Agencies' accomplishment of mission programs that may include specifications for rapid response or access. Network design efforts also attempt to create a robust network, one that is insensitive to failure. *Telecommunications Enterprise Network Design, Comprehensive Baseline Analysis: Task III, January 6, 1998*, focuses on identifying the types of circuits and the use of circuits for the current USDA Data Networks. Although recommendations are made regarding optimizing specific agency circuits, the final optimization recommendations depend on the current study's assessment of application level traffic.

2.0 Methods

2.1 USDA Data Network Application Traffic.

2.1.1 'NA Analyzer' Deployment Plan

An important network analysis tool for assessment of network application traffic is Network Associate's Sniffer Network Analyzer (NA Analyzer)[1]. NA Analyzers are installed at key locations on the USDA Data Network either in stand-alone or distributed configurations. The data obtained from the NA Analyzers is instrumental in building traffic model demands to be used in the analysis of the current USDA Data Networks. In addition, traffic models are required for modeling the Telecommunications Enterprise Network (TEN) design alternatives. For many reasons, including time and resources, it is impossible to collect 100% of a data networks' traffic. Therefore, proper placement of the NA Analyzers is crucial for developing accurate representation of network traffic and trends.

There are currently over a dozen NA Analyzers strategically located to capture LAN and WAN traffic on the USDA Data Networks. The traffic data being collected is generated from two major sources:

- Internet access network (Internet, Kansas City-NITC, New Orleans-NFC, Washington-HQ)
- Agency networks

[1] Sniffer Network Analyzer is a trademark of the Network General Corporation

The Internet access network data is captured from the following locations:

- Fort Collins - 1 NA Analyzer
- Kansas City - 2 NA Analyzers
- Washington DC (Headquarters) - 2 NA Analyzers

The USDA Agency specific traffic being captured in various locations are:

- AMS – 1 NA Analyzer
- APHIS - 1 NA Analyzer
- Forest Service - 5 NA Analyzers
- NRCS, ARS, LAN/WAN/Voice, and portions of other Agencies' traffic from Internet Access Network sniffers
- NASS – Agency managed sniffers (currently no data available)
- RD – Agency managed sniffers (currently no data available)

For some agency managed sniffers, there is currently no data available. This is due to resource constraints that have prevented agency personnel from providing sniffer files to OCIO. Where this data was not available, survey data from the agency is used to determine the traffic model for the agency. This process is described in section 2.2.

2.1.2 Traffic Data Collection Methodology

The basic function of the NA Analyzer is to collect a selected set of traffic [filtered via Internet protocol (IP) addresses] during a pre-determined time period. The NA Analyzer is capable of analyzing over 120 protocols, on Ethernet, Token Ring and Wide Area Networks [e.g. Dedicated Transmission Service (DTS), Frame Relay (FR), etc.] To obtain the most comprehensive and accurate data possible, within the project constraints, several primary sampling methods are used. However, the process is kept flexible enough to allow adjustments for special circumstances.

For the three sampling schedules, all data is collected during peak traffic periods. Peak periods are identified from network utilization statistics collected from USDA networks. This data was sampled not only over normal business hours, but also over hours at night. The data indicated that peak utilization times occur during normal business hours. Sampling methods are as follows:

- <u>Two samples per day:</u> With this method the first 2-hour capture occurs at peak periods between 7:00 AM and 12:00 PM local time. The second 2-hour capture is in the afternoon at peak periods between

12:30 PM to 6:00 PM local time. These samples provide a good cross representation of traffic within a USDA workday.

- Three samples per day: This sampling method uses one 2-hour and two 1.5-hour captures. Different IP filtering is done for each capture. This method provides a broad cross-section of the various network traffic patterns during different periods of the day.
- Four samples per day: This sampling method uses four 1-hour captures per day with capture times between 7:00 AM to 6:00 PM local time. This method is used to provide manageable data for very high traffic volume.

NA Analyzers save most of the information as ASCII files in the Expert Analyzer database. All data in the ASCII files is in comma separated value (CSV) format that allows exporting of the file information to spreadsheet, database applications and also to the NetMaker XA®[2] tool for modeling.

Among the different sets of transaction data that can be obtained from the NA Analyzers, the only fields retained are those NetMaker XA requires to generate and process demands. The following information is used:

- network protocol identification
- transaction source and destination addresses
- application type
- transaction start and stop time
- transaction size in bytes and packets for forward and return directions.

Table 1 shows examples of two transactions captured by a probe and reflect the information (fields) that are being used by NetMaker XA for traffic modeling.

Network Protocol	Source Address	Destination Address	App. Type	Start Time	Stop Time	Byte Forward	Byte Return	Packets Forward	Packets Return
IP	xx.xx.xx	yy.yy.yy	www -http	1/2/98 15:31	1/2/98 15:31	40,870	8,468	74	146
IP	zz.zz.zz	nn.nn.nn	TCP	1/2/98 15:31	1/2/98 15:31	227,100	454,200	3,785	7,570

Table 1 Example of NA Analyzer Data

It should be noted that all sniffer data is collected and stored according to departmental policy. Specifically, only packet header information is sampled from the network (i.e. packet content is not collected). Only authorized personnel are permitted access to the sniffer data. Network

[2] NetMaker XA is a registered trademark of MakeSystems, Inc.

security is in place for the Network Engineering Division LAN and the data is stored on secure servers. All hard copy data is destroyed when analysis is completed.

2.2 Traffic Modeling

Network traffic modeling provides the capability to evaluate network behavior using network development tools (e.g. NetMaker XA). The simulation parameters are derived from models representing the overall USDA Data Networks traffic in terms of traffic type (application), and traffic load (volume). Traffic models are used in the following USDA Network Design Process sub-tasks:

- baseline survivability assessment
- baseline performance assessment
- new design alternative(s) optimization
- new design alternative(s) survivability assessment
- new design alternatives(s) performance assessment

The USDA Data Networks traffic models are created from a set of components that include traffic demands, service points, and LAN sizing (Section 2.2.4).

2.2.1 Traffic Demands

Traffic demand is the modeling "unit" that forms the basis for application level modeling with NetMaker XA. Traffic demand represents a **request for** a particular **service** between two points on a network. A traffic demand is identified from actual network traffic samples captured with an NA Analyzer (see Section 2.0) and is characterized by its:

- network protocol (always Internet Protocol (IP));
- source and destination addresses;
- application type;
- sampling start and stop time; and
- size specified in bytes and packets for both transmission directions

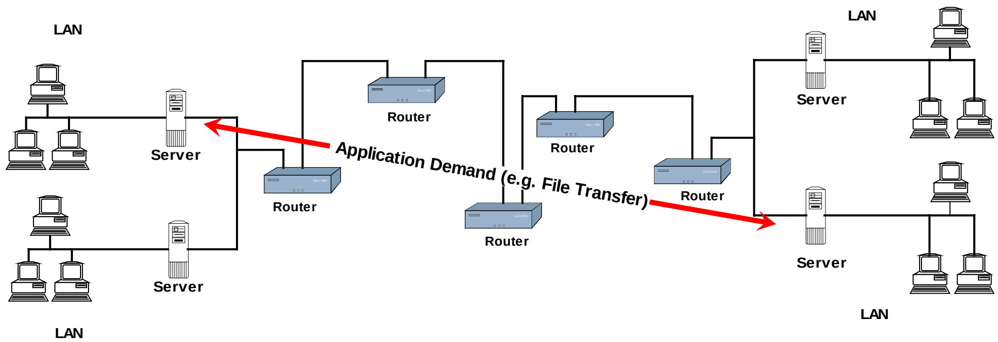


Figure 1 Relationship between Physical Topology and Demands

Traffic demands provide a reasonably accurate model of the data flow on existing or new networks. As a request for service, traffic demand is independent of WAN topology and equipment (Fig. 1).

2.2.2 Service Points

Service points are key elements to establishing normalized application traffic demand values. A service point is a LAN featuring a network service accessed by multiple user LANs (Fig. 2). Services include Internet, Web servers, Mail servers, mainframe processing centers, domain name service and other value-added services. Monitoring service points, therefore, essentially measures application type traffic.

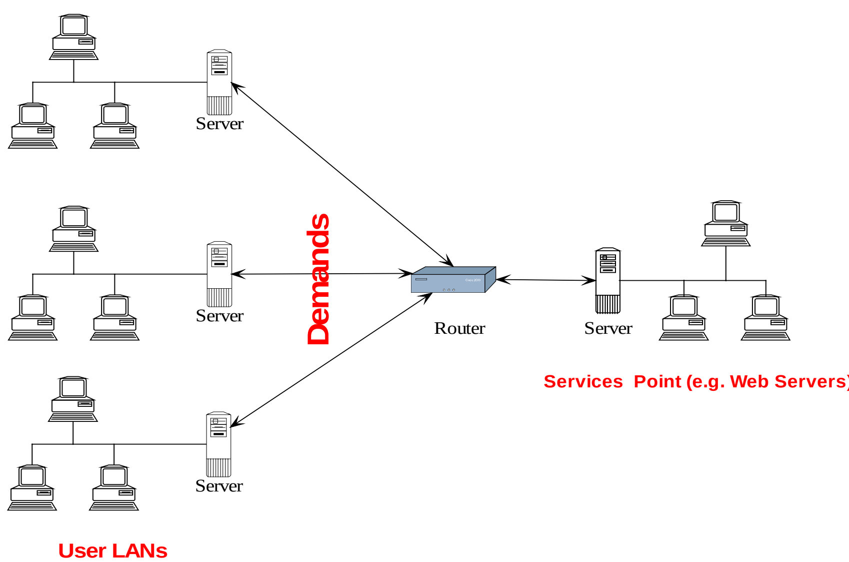


Figure 2 Service Point Definition

Service points for USDA data networks are identified from two sources:

- Application level data sampled from the USDA networks (Section 2.1.2) is derived from the NA Analyzers and sorted by LAN and Agency. The service point LANs are selected based on the high frequency of individual interconnections.
- Survey information from individual USDA Agency technical points of contact provides locations of known service points. The information is then correlated with the appropriate LANs in the network model.

The list of service points from the two sources is then consolidated (service points supporting the same demand types within the same geographical region are combined) and associated with their respective Agency. The list is then compared with the discovered network model. If required, the network model is adjusted (service point LAN placed at the appropriate model's location). Only the high traffic volume sites are currently used for modeling. As additional information becomes available, the service points list is updated accordingly.

2.2.3 USDA LAN Sizing

The USDA Agency population, or "head-count", is used to size and scale application level traffic demands for each modeled LAN. The "head count" per LAN represents the sizing parameter for network modeling. The sizing parameter is used to "load" the network model with the appropriate traffic volume -number of traffic demand instances generated by an individual.

Establishing the LAN sizing parameter requires the use of third party information (Table 2A). The USDA National Finance Center (NFC), Office of Personnel (New Orleans) provided the number of USDA employees categorized by state, city, and agency. Since NFC data is limited to US based permanent full-time (PFT) employees, LAN sizing values reflect the same limitations. Despite limitations, initial LAN sizing is sufficient for modeling purposes. As more complete information becomes available, LAN sizing figures are adjusted.

Table 2A

Office of Personnel			
State	**City**	**Agency**	**# of PFT**
AZ	Phoenix	A	57
AZ	Phoenix	B	38
MT	Plains	B	58

Table 2B

"Discovered" LAN Sizing					
LAN Adrs.	**Agency**	**Street**	**City**	**State**	**LAN-Size**
xxx.yyy.nn	A	S. Gila Avenue	Phoenix	AZ	28
xxx.yyy.pp	A	Cactus Road	Phoenix	AZ	29
zzz.yyy.nn	B	Dos Amigos St.	Phoenix	AZ	38
ttt.zzz,mm	B	Mountain Street	Plains	MT	29
ttt.zzz.xxx	B	Big Pine Avenue	Plains	MT	29

Table 2 A & B Example of LAN Sizing Data

Each Agency employee "head count" within a city is associated with the network-discovered information (LAN, agency, city, and state). If an Agency within the same city has multiple LANs, the employee 'head count' is divided equally between the Agency's LANs (Table 2B).

2.2.4 Building Traffic Models

Traffic models are crucial to the simulation of telecommunications networks. Models represent the current or anticipated application traffic to be supported by the network. It is essential that traffic models represent, as close as possible, real traffic in terms of frame/packet size, rate and volume.

The Traffic Model Generation Process (Fig. 3) uses the generated service point list (Section 2.2.2), the Agencies' LAN sizing (Section 2.2.3), and sampled application traffic of USDA Data Networks (Section 2.0), to produce a set of normalized traffic demands representing a generalized view of the USDA Network traffic. The traffic demands generated are used to analyze the current USDA data networks. They also are used as the fundamental elements for the design of the TEN alternatives.

A set of basic demand types is identified from the list of service points as described in Section 2.2.2. When actual sampled data associated with a particular demand type is available, the packet size, transaction size, and transaction rate attributes are calculated. The NETMAKER XA simulation process requires these attributes. When there is no sampled data associated with a particular demand type, attributes of a similar demand type is used. The equations below are used to calculate the attributes required to create traffic demands in NetMaker XA. The initial packet and transaction attributes are extracted from actual traffic samples obtained from the data probes.

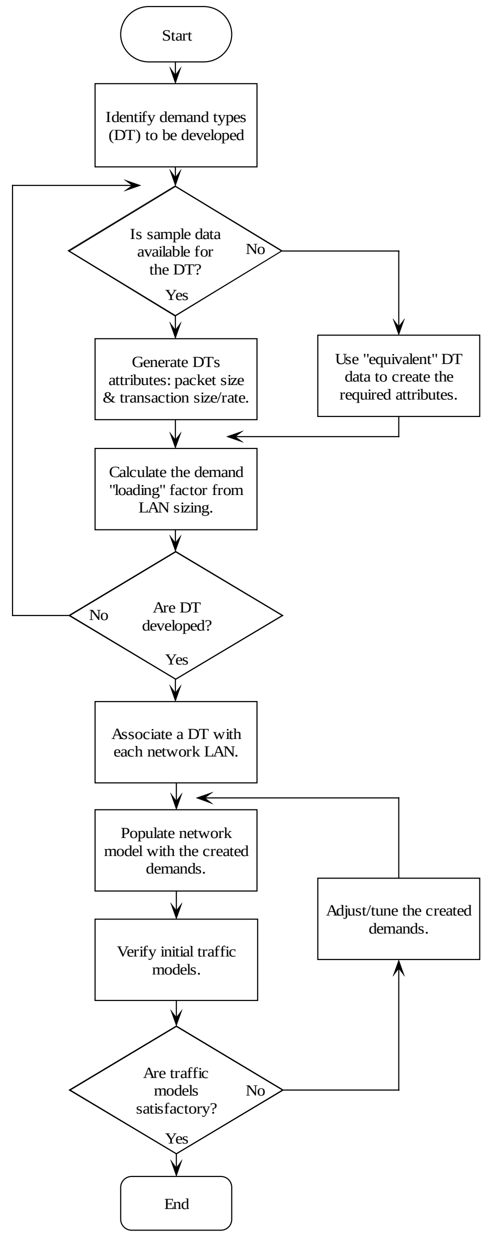


Figure 3 Traffic Model Generation Process

$$Packet\ Rate\ Forward_i = \frac{Transaction\ Rate_i \bullet Transaction\ Size\ Forward_i}{Packet\ Size\ Forward_i}$$

$$Packet\ Rate\ Forward_i = \frac{Transaction\ Rate_i \bullet Transaction\ Size\ Forward_i}{Packet\ Size\ Forward_i}$$

$$Packet\ Rate\ \mathrm{Re}verse_i = \frac{Transaction\ Rate_i \bullet Transaction\ Size\ \mathrm{Re}verse_i}{Packet\ Size\ \mathrm{Re}verse_i}$$

$$Population\ Packet\ Size\ Forward_i = \frac{\sum_i^n (Packet\ Rate\ Forward \bullet Packet\ Size\ Forward)}{\sum_i^n (Packet\ Rate\ Forward)}$$

$$Population\ Packet\ Size\ \mathrm{Re}verse_i = \frac{\sum_i^n (Packet\ Rate\ \mathrm{Re}verse \bullet Packet\ Size\ \mathrm{Re}verse)}{\sum_i^n (Packet\ Rate\ \mathrm{Re}verse)}$$

$$Population\ Transaction\ Size\ Forward = \frac{\sum_i^n (Transaction\ Size\ Forward \bullet TransactionRate)}{\sum_i^n (Transaction\ Rate)}$$

$$Population\ Transaction\ Size\ \mathrm{Re}verse = \frac{\sum_i^n (Transaction\ Size\ \mathrm{Re}verse \bullet TransactionRate)}{\sum_i^n (Transaction\ Rate)}$$

$$Population\ Trasaction\ Rate = \sum \frac{Transaction\ Rate_i}{n}$$

To "load" the network model with a reasonable traffic volume, there is a NetMaker XA "multiplier" to set the number of times the same demand type transaction should occur. In the TEN design process, the NetMaker XA multiplier is derived from the LAN-size (Section 2.2.3). To define the NetMaker XA multiplier, the demand type individual transaction must be determined.

$$Unit\ Transaction\ Rate = \frac{Population\ Transaction\ Rate}{\sum_i^n (LAN - size)}$$

The Unit Transaction Rate denominator, Σ (LAN-size)$_n$, represents the total number of "individuals" participating in the generation of a service point traffic demand. The "count" number used to simulate the traffic from a particular LAN.

For each USDA Agency, organizational LANs are associated with the identified service points. Providing demand attributes and specifying the “count” as described above generates the traffic demands for each LAN. Those demands are imported by NetMaker XA network model for simulation.

The initial set of traffic demands is subjected to a verification process prior to confirmation as "USDA Traffic Model". The network loaded with the initial set of demands is simulated and a utilization report is created for each circuit. The utilization report is compared with data reported in *Telecommunications Enterprise Network Design, Development of Network Capacity Design: Task IV Report, March 13, 1998.* Unacceptable differences between the two reports necessitate re-tuning of the affected traffic model (e.g. variation of the ‘count’ parameter).

2.3 USDA Data Networks Performance

The performance analysis of the USDA Data Networks uses traffic demand models, derived from actual USDA application traffic (see Section 3.0 above), applied to the January 30, 1998 USDA Data Networks topology "Discovery". The traffic demands are used in the simulation of the networks to derive a general view of the current performance of the networks. The simulation of the networks uses NetMaker XA’s ‘Planner’ simulation process. Two major steps are performed during this process:

- routing of each of the traffic demands using the routing protocol specified for each router in the networks - Open Shortest Path First (OSPF) routing protocol is used for all routers during the simulation; and
- performance calculation and reports generation.

The performance analysis of the USDA Data Networks is divided into two topics:

- traffic demand delays between source and destination
- utilization of the WAN links based on the application traffic. WAN Links include DTS and Frame Relay accesses.

It is important to remember that the ‘simulation’ of the network is based on the traffic models. Those models were created from **actual USDA Data Networks application traffic** and were made to represent, as closely as possible, the traffic on the existing USDA Data Network.

2.4 USDA Data Network Survivability

Survivability of a network is defined as its ability to support the network traffic when one or more nodes and/or links defined for that network fail. As for the performance analysis of the USDA networks, the survivability analysis uses the application traffic model that has been created for network simulation purposes (see Section 3.0). NetMaker XA uses its "Analyzer" and "Planner" modules combined to simulate realistic network failure situations. The results of the analysis are based on whether or not traffic demands can reach their destination because of certain failure conditions being simulated.

2.4.1 Process

The "analyzer" module of NetMaker XA performs the network survivability assessment by:

- automatically and systematically simulating failing groups of objects such as WAN links and executes simulated alternative design for each configuration of active network objects;

- calculating aggregate results from the suite of failures investigated such as the Network Survivability Index (NSI) values;

- generating analysis results using the tool's internal reporting process.

Due to the large number of nodes, links, and traffic demands, it is not possible to fail all network objects in the time frame given for the project. Because of this, two adjustments were made for the survivability analysis. The first adjustment was to simulate only WAN link (DTS and FR Access) and LAN failures. Routers were not failed. In addition, only the Single Failure Simulation mode is used. In this mode, NetMaker XA will simulate a systematical failure of all the LAN, DTS and Frame Relay Access links, one at a time. The second adjustment was to reduce the number of traffic demands modeled during the simulation. Only Internet traffic demands were simulated. The Internet traffic was selected because these demands were common to all agencies, and they followed similar routing paths as the other modeled traffic demands. This adjustment will be duplicated when the design alternatives are analyzed for network survivability.

3.0 Results

3.1 Definition of Network Application Terms

Internet	Traffic exiting from the USDA Network to the Web.
intranet	Traffic that stays on the USDA network.
count	The number of conversations (or sessions) that occurred during the measuring period. A conversation is defined as a transaction between two network devices. For example, a file transfer is considered one conversation.
bytes	The volume of traffic that occurred during the measuring period expressed as the total number of bytes.
HTTP	Hypertext Transfer Protocol - The protocol used by Web browsers and Web servers to transfer files, such as text and graphic files. This is traffic associated with Web browsing.
TCP	Transmission Control Protocol - Network traffic that could not be identified by well-known port numbers
NetBIOS	Network Basic Input/Output System - API used by applications on an IBM LAN to request services from lower-level network processes. Services might include session establishment and termination, or information transfer.
SMTP	Simple Mail Transfer Protocol - Traffic associated with transferring E-mail between devices.
UDP	User Datagram Protocol - UDP traffic that could not be identified by well-known port numbers.
POP3	Post Office Protocol version 3 - Protocol that client e-mail applications use to retrieve e-mail from an e-mail server.
Telnet	Standard terminal emulation protocol in the TCP/IP protocol stack. Telnet is used for remote terminal connection, enabling users to log in to remote systems and use resources as if they were connected to a local system.

FTP	File Transfer Protocol – A network application protocol, part of the TCP/IP protocol stack, used for transferring files between network nodes.
other	Other TCP/IP traffic associated with well-known port numbers such as Domain Name System (DNS), X-terminals (X11), Trivial File Transfer Protocol (TFTP), etc.

3.2 Application Traffic Analysis

Application traffic information is a summary of the data sampled from the USDA network. This information is obtained from the strategically located NA Analyzers using the methodology described in Section 2.1.2. Because it is not feasible to sample data at every point in the USDA network, the network application traffic analysis technically represents only the portion of the network that was sampled. However, because the data was collected at backbone concentration points, it typifies traffic on the USDA network.

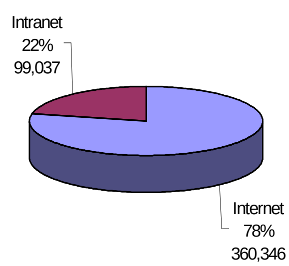


Figure 4 Distribution of Traffic Sessions Between the Internet and Intranet

Figures 4 and 5 show the distribution of Internet and intranet traffic measured on the Existing USDA Data Networks. Figure 4 shows traffic by count, representing the number of conversations (sessions) that occurred during the sampling period. Figure 5 represents traffic volume, expressed as total number of bytes, during the sampling period expressed as the total number of bytes.

NOTE: Figures 4 through 11 present the data both as a calculated percentage as well as the actual "count" or "bytes" measured.

The volume of **modeled** USDA network traffic on the Internet is 19.1% (262,211,076 bytes) while the volume of Intranet traffic is significantly larger at 80.9% (1,110,151,013 bytes). The volume of **measured** USDA network traffic on the Internet is 69% while the volume of Intranet traffic is significantly smaller at 31% (Fig. 5). (Count data is not available for the modeled network because the model is based on LAN to LAN volumes, and not individual conversations.)

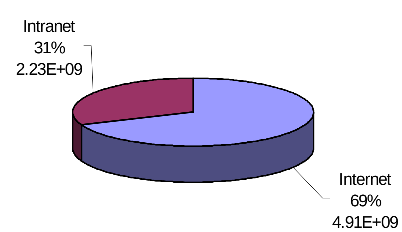


Figure 5 Traffic Volume Distribution Between Internet and Intranet

Comparing the traffic distribution for the **modeled** Networks with that of the **measured** Networks indicates a discrepancy in the Internet/intranet volume distribution. This discrepancy is due to the addition of survey data to the modeled application traffic. Survey data has the effect of adding traffic that is projected for the near future, but has not been deployed. For example, there was very little traffic measured on the existing network that went to NITC or NFC. However, survey data indicates that most agencies do or will communicate with NITC and NFC over the TEN. By modeling this survey data, the amount of Intranet traffic on the modeled network is increased.

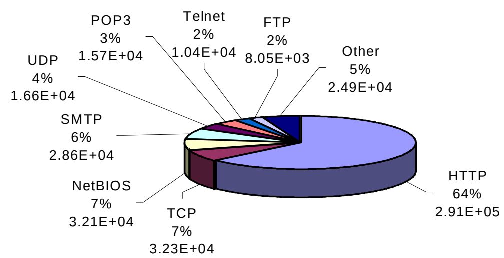


Figure 6 Network Traffic - Combined Summary: Application Sessions

Figures 6 and 7 show the types of network applications in use and their distribution on the Internet and intranet. The network application types are described in Section 3.1. Figures 8 and 9 show the types of network applications measured going to the Internet. Figures 10 and 11 show the types of network applications measured staying on the USDA intranet.

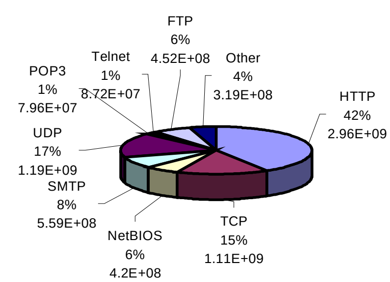


Figure 7 Network Traffic - Combined Summary: Application Volume

The Application Level Traffic analysis is the last task of the "baseline" modeling of the existing USDA Data Networks (see USDA Network Design Process). The results of the analysis presented in this document are consistent with the results presented in previous baseline documents. In particular, the results related to overall network utilization assessment and topology characteristics have been the same throughout the process.

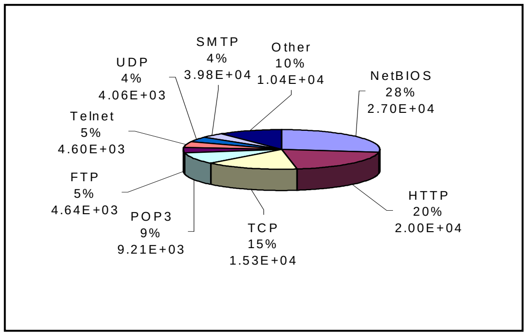
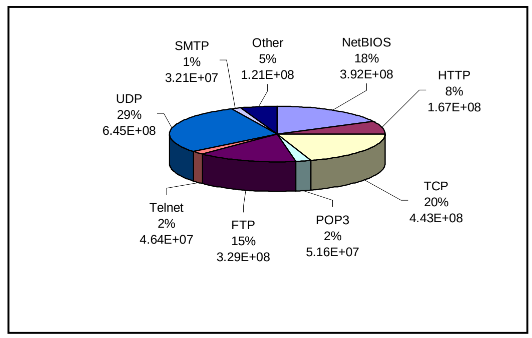


Figure 8 Intranet Summary: Application Sessions – The measurement of the types of network applications initiated during the sampling period as intranet traffic indicates a fairly broad distribution. Nevertheless, NetBIOS (28%) and HTTP (20%) represent nearly half the application sessions.

Figure 9 Intranet Summary: Volume – The distribution of the various network applications use d on the intranet is primarily divided between UDP (29%), NetBIOS (18%), and FTP (15%).



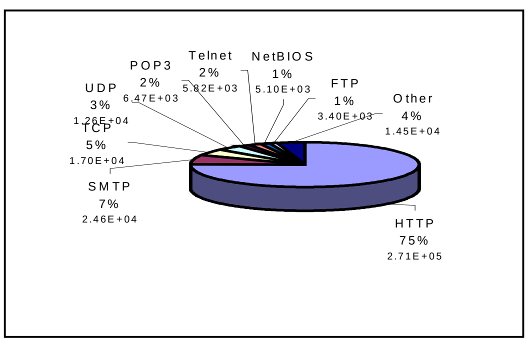
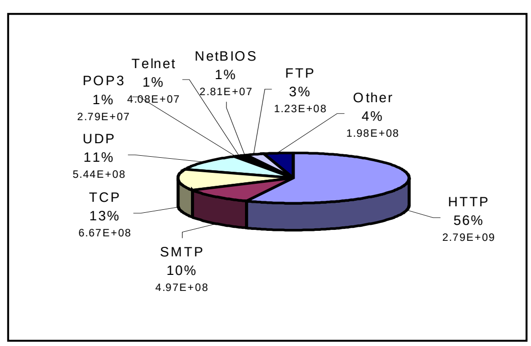


Figure 10 Internet Summary: Application Sessions – Of the nine categories of network applications measured, HTTP is the predominate *75%) application initiated by USDA Agencies.

Figure 11 Internet Summary: Application Volume - During the sampling period, 56% of the USDA traffic on the Internet is Web browsing.



3.3 Traffic Modeling

Traffic model simulation of the USDA Data Networks (Section 2.2.4) generated the data describing the existing Network's operating characteristics. Although the results are calculated, they are useful for providing a baseline description of the existing USDA Data Networks for comparison with design alternatives.

By constructing traffic demands using the process described in section 2.2, a total of 10,291 demands were modeled. The service points used to construct the traffic model are presented in Table 3. The simulation of the existing network traffic yielded an average WAN link utilization of 4.783%. This is compared to a measured average WAN link utilization of 4.378%, and an artificial high WAN link utilization of 8.215% (These numbers were developed as a part of *Telecommunications Enterprise Network Design: Network Level Traffic Study of USDA Data Networks -Task II, December 16, 1997*). It was not possible to reproduce the artificial high WAN link utilization because there was insufficient capacity on existing network "backbone" links to handle the simultaneous traffic demands needed to load the high day traffic data. Because of this, the traffic analysis was determined with the average utilization data. The same set of traffic demands will be used when comparing new designs to the existing network.

Agency	Service	Notes
AMS	• Internet • AMS-HQ • DC-Server • NFC • NITC	Internet Traffic Washington Headquarters LAN Washington DMZ server LAN National Finance Center, New Orleans National Information Technology Center, Kansas City
APHIS	• Internet • APHIS-FC1 • APHIS-FC2 • APHIS-HQ1 • APHIS-HQ2 • APHIS-NJ	Internet Traffic Fort Collins Server LAN Fort Collins Server LAN Washington Server LAN Washington Server LAN New Jersey Server LAN
ARS	• Internet • NFC • ARS-HQ • ARS-FC	Internet Traffic National Finance Center, New Orleans Washington Headquarters LAN Fort Collins Server LAN
CSREES	• Internet	Internet Traffic
FAS	• Internet • NFC • NITC • FAS-HQ	Internet Traffic National Finance Center, New Orleans National Information Technology Center, Kansas City Washington Headquarters LAN
FNS	• Internet • FNS-HQ	Internet Traffic Washington Headquarters LAN

Agency	Service	Notes
FS	• Internet • FS-HQ • FS-SERVER1 • FS-SERVER2 • FS-SERVER3 • FS-SERVER4 • FS-SERVER5 • FS-SERVER6 • FS-SERVER7 • FS-SERVER8 • FS-SERVER9	Internet Traffic Washington Headquarters LAN Regional Server Regional Server Regional Server Regional Server Regional Server Headquarters Server Regional Server Regional Server Regional Server
FSA	• Internet • FSA-HQ • FSA-KC • NFC • NITC	Internet Traffic Washington Headquarters LAN Kansas City Server LAN National Finance Center, New Orleans National Information Technology Center, Kansas City
FSIS	• Internet • FSIS-HQ • NFC • NITC	Internet Traffic Washington Headquarters LAN National Finance Center, New Orleans National Information Technology Center, Kansas City
GIPSA	• Internet • GIPSA-HQ • NFC	Internet Traffic Washington Headquarters LAN National Finance Center, New Orleans
NASS	• Internet • NASS-HQ • NASS-LM • NFC • NITC	Internet Traffic Washington Headquarters LAN Florida Server LAN National Finance Center, New Orleans National Information Technology Center, Kansas City
NRCS	• Internet • NRCS-FC • NITC • DC-SERVER	Internet Traffic Fort Collins Server LAN National Information Technology Center, Kansas City Washington DMZ server LAN
OCE	• Internet • NFC • NITC	Internet Traffic National Finance Center, New Orleans National Information Technology Center, Kansas City
OCFO/NFC	• Internet	Internet Traffic
OCIO	• Internet • DC-SERVER • NITC	Internet Traffic Washington DMZ server LAN National Information Technology Center, Kansas City
RD	• Internet • RD-SL • RD-HQ • DC-SERVER • NITC	Internet Traffic Saint Louis server LAN Washington Headquarters LAN Washington DMZ server LAN National Information Technology Center, Kansas City
RMA	• Internet • RMA-KC • DC-SERVER • NFC • NITC	Internet Traffic Kansas City Server LAN Washington DMZ server LAN National Finance Center, New Orleans National Information Technology Center, Kansas City

Table 3 Types of Network Applications by USDA Agency

3.4 Performance

3.4.1 Delay Assessment

Delay metrics describe the performance of the existing networks and provide a reference to assess delay of TEN models in the development of the alternative TEN designs.

Although many USDA Data Networks are underutilized as discussed in *Telecommunications Enterprise Network Design, Comprehensive Baseline Analysis – Task 3, January 6, 1998*, some were discovered to have over-utilized circuits on critical paths. Over-utilization was determined to be as much as 280% of the theoretical capacity in some circuits. This over-utilization causes transmission bottlenecks, which result in infinite delays when attempting to do a network simulation. In the simulated network, 81.6% of the modeled traffic demands failed due to infinite delays. With network modeling, infinite delays cause simulation to fail, whereas with actual networks, the demands do not fail but result in loss of data packets and increased round-trip delays. In order to allow network modeling without failed demands, it was necessary to increase bandwidth at bottlenecks to force less than 100% utilization. This artificial condition imposed on the model makes it more difficult to compare existing networks and new designs because the simulated round-trip delay is less than the actual delay. The simulated round trip delay averages will be used for the comparison between new design alternatives.

The major components contributing in the calculation of the simulated network delays are queuing delay, transmission time, and propagation delay:

- NetMaker XA uses the Traffic Demand arrival rate, mean packet size, and mean service rate characteristics in the modeling of router queues. The Pollaczeck-Khinchin mean value formula for a M/G/1 queue is used to determine the mean length of the queue in the router's interface.

- The transmission time (propagation delay) is determined based on the average packet size and link speed.

- NetMaker XA calculates the propagation delay based on the length of links.

The delay assessment shows that in the combined USDA Data networks, after correction to eliminate infinite delays, **the average traffic demand round trip delay is 42.5 milliseconds**. This number is calculated by averaging round trip delay of the 10,291 traffic demands model. The average round trip delay per demand is the sum of the average forward and return delays (per demand). As a comparison benchmark, commercial Frame Relay service providers have performance guarantees varying from 70 to 300 msec for delays between the origin and destination POPs of a PVC.

The USDA networks traffic demands delay modeled **ranges from .5 msec to 356 msec**. The latter being the maximum demand delay found among the entire set of traffic demands. The distribution of percentage of demands within a given range of delay is presented in Figure 12.

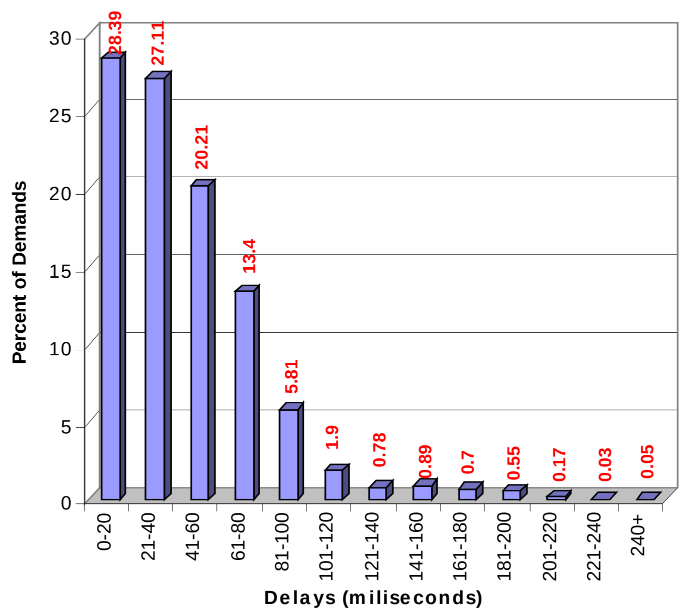


Figure 12 Percentage of Demands with Delays

The **number of round-trip router "hops"** that a traffic demand encounters between its source and destination contributes to the traffic demand delay. With respect to this document and specific process, a "hop" is defined as an intermediate router device in a string of connections linking two network devices (Fig 13). Theoretically, fewer hops between the source and destination means less delay. However, other factors such as network congestion and "busy" routers may increase delay.

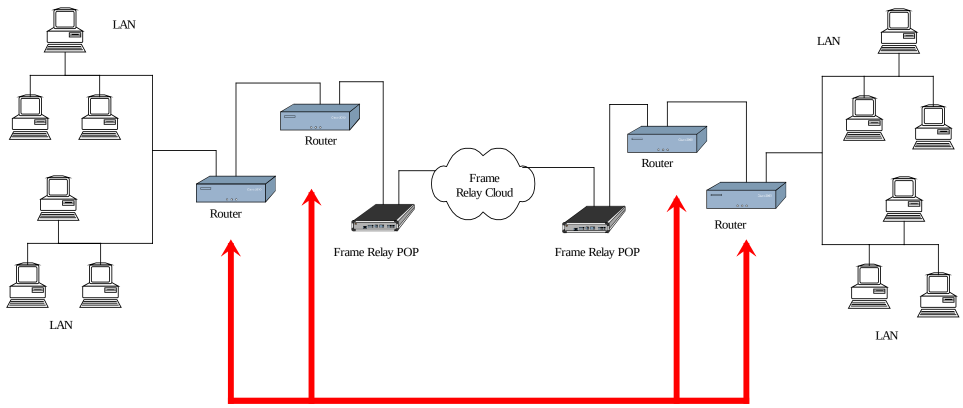


Figure 13 Router "Hops" - A router hop is defined as an intermediate router device in a string of connections linking two network devices. In the diagram, four router hops are indicated.

Number of hops metrics—in particular the average number of hops—may be viewed as a network design efficiency factor. The current USDA network hop count metrics are used as established baseline figures to evaluate efficiency of the TEN alternative designs. Figure 14 illustrates the distribution of the number of round-trip hops used by traffic demands in the network model. Some useful statistics are derived from this data.

- The most number of round-trip router hops used by a traffic demand is 24

- The least number of round-trip router hops used by a traffic demand is 2

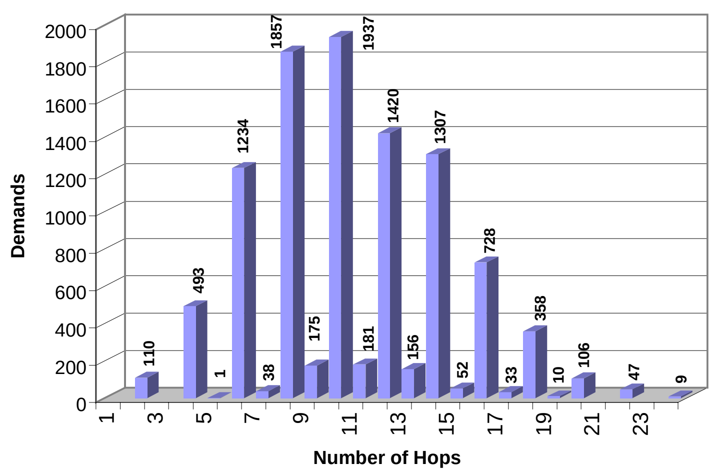


Figure 14 The Number of Round-trip Router "Hops" Used by Traffic Demands

- The average number of round-trip router hops used by the USDA Data Networks traffic demands is 10.5

The **number of round-trip network "links"** that a traffic demand encounters between its source and destination contributes to the traffic demand delay. With respect to this document and specific process, a "link" is defined as an intermediate connection in a string of connections between two devices (Fig. 15). Links include items such as FR links, DTS links, LAN links, FR clouds, FR access, and FR pops. Like the number of router hops, fewer links between the source and destination generally means less delay. The majority of the end-to-end routes in USDA Data Networks include LEC and FTS 2000 "clouds". Since the number of links within the service provider "clouds" is unknown, each instance of a "cloud" within the route path is considered as one link.

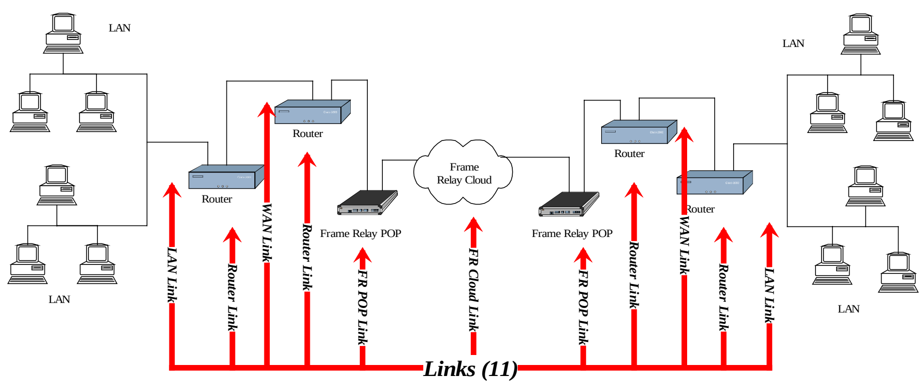


Figure 15 Round Trip Links - A link is defined as an intermediate connection in a string of connections between two network devices. The diagram indicates 11 links of 5 types.

Number of links metrics—in particular the average number of links—may be viewed as a network design efficiency factor. The current USDA network link count metrics are used as established baseline figures to evaluate efficiency of the TEN alternative designs. Figure 16 illustrates the number of links used by traffic demands in the network model. Statistics derived from this data include:

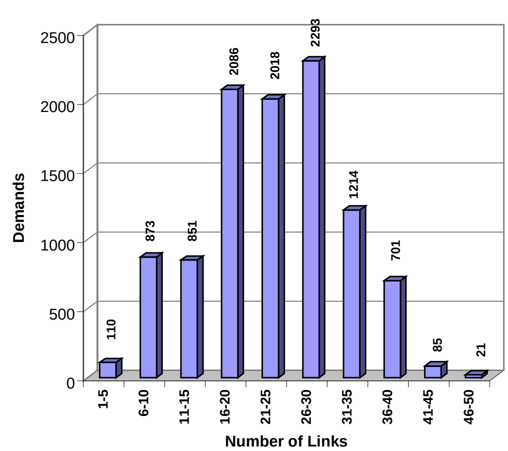


Figure 16 The Number of Round-trip Links Used by Traffic Demands

- The most number of round-trip network links used by a traffic demand is 46

- The least number of round-trip network links used by a traffic demand is 4
- The average number of round-trip network links used by the USDA Data Networks traffic demands is 23.2

3.4.2 WAN Link Utilization Based on Network Application Traffic

WAN link utilization variables are derived from actual application traffic used in simulating and describing traffic demands. Utilization and network delays are the major factors for evaluating network performance. While performance addresses network speed, utilization addresses volume and link capacity. The metrics provide information on the current networks and, in addition, provide the utilization reference for TEN development purposes (e.g. How big should a pipe be?).

The **total active WAN bandwidth** is the total amount of bandwidth in all WAN links for all networks. It includes WAN link forward / return bit rate and Frame Relay PVC Committed Rate (CR) or Burst Rate (BR). The total active WAN bandwidth is 36,392,942.000 Kbits/s. This value is the sum of the used and unused bandwidth specified below.

The **unused WAN bandwidth** is the amount of total bandwidth in all networks that is not used. The USDA networks unused bandwidth represents 99.899 % of the total active bandwidth, or 36,356,025.464 Kbit/s.

The **used WAN bandwidth** is the amount of total bandwidth in all networks that is used. The USDA networks used bandwidth represents 0.101% of the total active bandwidth, or 36,916.536 Kbit/s.

The **maximum WAN link utilization** represents the highest utilization for any one link within the USDA networks. Once over-utilized circuits were recapacitated to allow the model to work, the highest utilization is 92.003% of its theoretical capacity. Without adjustments for over-utilization, the maximum WAN link utilization is 284% of its theoretical capacity.

The **average WAN link utilization** represents the average of all the USDA WAN links utilization based on application traffic. The average is 4.783 % of the WAN links combined capacity.

3.5 Survivability

The results of the USDA Network Survivability analysis represent a measure of current USDA Data Networks robustness relative to link failure. Due to the large number of nodes, links, and traffic demands, it is not possible to fail all network objects in the time frame given for the application study deliverable. Because of this, adjustments were made for the survivability analysis. The first adjustment was to use single Failure Simulation mode for the analysis. In this mode, NetMaker XA will simulate a systematic failure of all the LAN, DTS and Frame Relay Access links, one at a time.

Another adjustment was to reduce the number of traffic demands modeled during the simulation. Only Internet traffic demands were simulated. The Internet traffic was selected because these demands were common to all agencies, and they followed similar routing paths as the other modeled traffic demands. This adjustment will be duplicated when the design alternatives are analyzed for network survivability.

The final adjustment was to simulate only WAN link (DTS and FR Access) and LAN failures. Routers were not failed. Given more time to complete the deliverable, it would be desirable to fail routers. Although router hardware failures are less frequent than link failures, network down time caused by router software configuration errors is common. A router failure analysis would give insight into this “human error” factor.

Survivability metrics also provide a very important reference to assess the robustness of the TEN design alternatives.

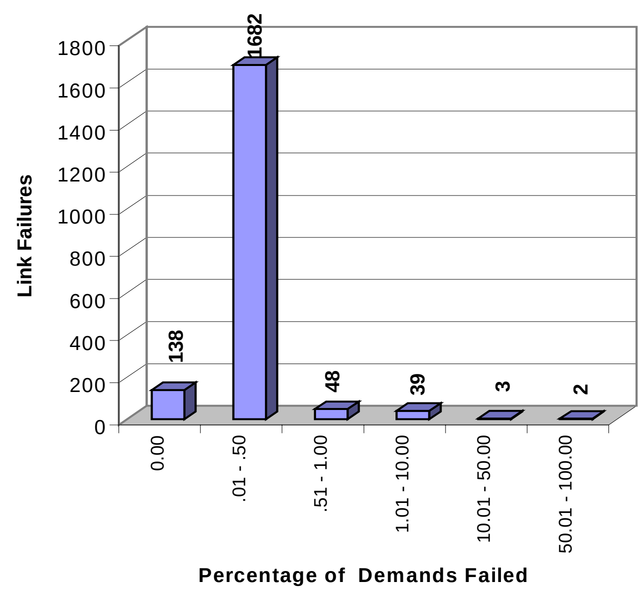


Figure 17 Percentage of Network Demands Affected by Link Failure

Figure 17 represents how the modeled Internet traffic of the USDA networks is affected by link failures. The diagram shows the distribution of failed traffic demands relative to link failures.

The link simulation as described above (Internet Traffic) provided the following results: The average percent network (Internet related) failure due to a single

LAN/WAN link failure is 0.241%. However, the maximum percent network failure caused by a single link failure is 99.72%. This is a direct result of network 'tree' structures where all access are concentrated to a single link access. In this particular case, while there was an alternative path to the internet traffic to access the cloud, the alternative path could not support additional traffic and totally collapsed under the 'weight' causing a near failure of the Internet related network.

4.0 Conclusions

- The baseline description developed in the four previous Network Engineering Division studies and the application model developed in this final task are sufficient to create TEN design alternatives.
- Internet is the predominant use for the existing network. Of the Internet traffic, Web browsing is the application most in use. From survey data, it appears that the percentage of Intranet traffic will increase. For example, there was very little traffic measured on the existing network that went to USDA data centers. However, survey data indicates that most agencies do or will communicate with the data centers over the TEN. The conclusion is that there is still traffic going to the data centers on non-IP based networks, such as dial X.25. This conclusion is supported by the continued increase in the X.25 cost in USDA billing data. Survey data indicates that this traffic will be moved to the IP based network, and will increase Intranet traffic on the TEN.

- Constraints on critical USDA Network paths affect performance, including delay, utilization and survivability. Predominant network designs on USDA networks are Minimum Spanning Tree (MST) and hub-and-spoke, which are characterized by bottlenecks. Analysis shows that on one USDA network with a design based on a MST structure, the simulated link utilization is 284%. This is in contrast to a USDA network with a design based on a semi-chordal Ring (SCR) structure, with a simulated utilization of 3%. A redesign of the USDA data networks will eliminate these bottlenecks and improve the overall performance of network applications.

  In general, the performance of the combined USDA data networks is relatively poor. Over-utilization was modeled to be as much as 280% of the theoretical capacity in some circuits. This over-utilization causes transmission bottlenecks that result in infinite delays when attempting to do a network simulation. In the simulated network, 81.6% of the modeled traffic demands failed due to infinite delays.

  Once the bottlenecks in the model are corrected by increasing bandwidth, the remainder of the networks is basically underutilized. In the modeled network, over 99% of the available bandwidth is not utilized. This gives the remainder of the USDA simulated networks very good performance ratings. The majority of the traffic demand delays are in the 0 to 60 msec range indicating very good network performance. As a comparison benchmark, commercial Frame Relay service providers have performance guarantees varying from 70 to 300 msec for delays between the origin and destination POPs of a PVC.

- One can derive, from this last study of the existing USDA networks, that improvements in performance, cost and survivability can be achieved within the TEN alternatives. The appropriate network design rules will improve survivability of the overall traffic and also will eliminate the bottlenecks; hence improve performance. Re-sizing of oversized link as well as sharing resources will contribute in reducing network cost.

- Survivability numbers indicate that on average single link failures cause a small number of Traffic Demands to fail. This is due to the large number of “feeder” circuits and LANs associated with small offices. However, there are bottlenecks in the networks that when failed cause large numbers of Traffic Demands to fail. This problem is related to the existing topology design discussed above.

Appendix - Refining Baseline Demands

Baseline demand data for the USDA Data Networks is measured with the NetMaker XA system. Data collected over a five-day period indicates that more then 3% of the WAN links appeared to exceed 100% link utilization. There are several possible explanations for the anomalous results. Incorrect MIB value for a WAN link access could cause the demand calculation to use a default T1 link rather than the actual 64K link. Incorrect demand data could also result from MIB traffic counters being reset due to scheduled maintenance or any other power interruption. Finally, incorrect demand data could result from a router or link malfunctioning between captures of MIB traffic counter values. Regardless of the origin of the high utilization values, it is important for analysis accuracy to use only correct demand data.

The high demand data may be viewed as either real or an erroneous outlier. In order to avoid skewed data, outlier data needs to be identified and eliminated from the demand calculation. Using the normal distribution theorem, it is possible to distinguish outlier data from legitimate high value data. Assuming the demand data is normally distributed, the normal distribution theorem provides the means to statistically distinguish between real and erroneous data (Fig. 18). For normally distributed data, 68% of sample observations occur within plus or minus one standard deviation of the mean (2 sigma). Similarly, 95% and 99% (4 and 6 sigma) of sample observations occur within plus or minus two and three, respectively, standard deviations of the mean. The traffic demand calculations used in the current study represent the four-sigma level of acceptance.

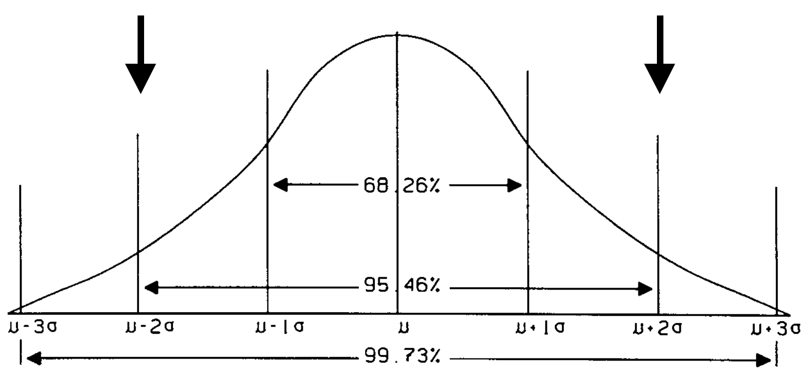


Figure 18 Normal Sample Distribution

The four-sigma level of acceptance means original observations were evaluated against the original mean plus or minus two standard deviations. If the observation falls outside this range, it was considered an outlier and eliminated from further analysis. The refined data set was then analyzed and the high hour was selected for each circuit. These type 1 demands were loaded into NetMaker XA. Planner was executed to simulate these demands. The resulting network average utilization was compared with the baseline network average utilization reported in *Telecommunications Enterprise Network Design, Development of Networks Capacity Design – Task IV, March 13, 1998.* The average utilization dropped from 14.977% to 12.243% even though more demands were used and more bytes were transferred. This would indicate that if four-sigma acceptance were applied to the baseline demands reported in Task IV, an even larger decrease in utilization would occur.

This would better correlate the baseline demands against the application demands in the current application demand analysis. The manual throttling of the application demands would then explain the primary differences in network utilization by the analysis to fit within the limitation of the NetMaker XA simulator. (Manual throttling is the restriction of a circuit's utilization to less than 100% by manipulating the transaction rate to insure that all demands transverse the network.)